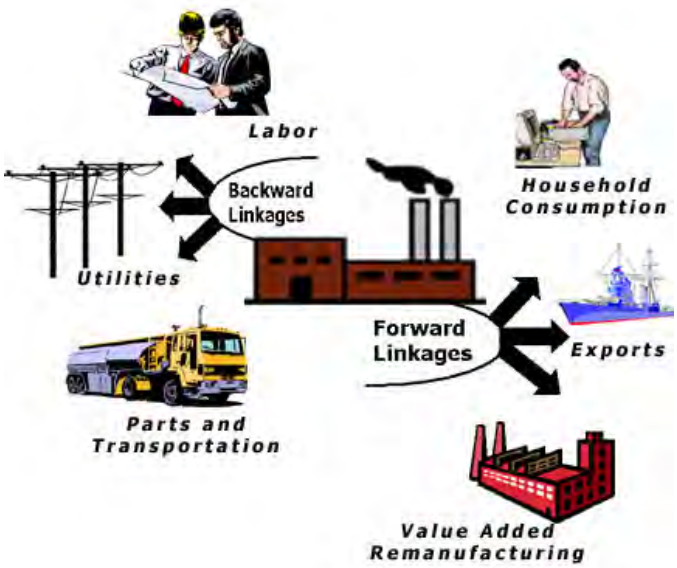
Missouri Industry Clusters:
Information Technology



INDUSTRY CLUSTER ANALYSIS

Report ICA-0601-1
June 2001

MISSOURI
RESEARCH AND PLANNING
DEPARTMENT of ECONOMIC DEVELOPMENT

Additional information is available on-line at:
http://www.MissouriEconomy.org

Missouri Industry Clusters:
Information Technology

INDUSTRY CLUSTER ANALYSIS

Report ICA-0601-1
June 2001

Analysis and reporting by David J. Peters, Planner.
Maps by Zachary Johnson, GIS Analyst.

I. Overview

Recently, there has been an interest in determining input and output industry clusters for a given economic sector - generally called an industry value-chain (Bergman and Feser 1999; Feser 1998). This interest has coincided with a focus on industry competitiveness and comparative advantage. In essence, each industry's competitive position depends on one or several supporting industries or institutions. This interdependence between an industry's suppliers and consumers is key to the success of a given industry. Industry cluster analysis views the development of supporting industries as vital to the health and growth of a given industry. Although industries can be clustered along labor and knowledge requirements, clustering along value-chains is more informative. Value-chains are detailed inter-industry transactions, which include sales to and purchases from every other sector.

Given that information technology (IT) is a targeted industry in Missouri, this analysis determines the value-chain industry clusters for the IT sector in Missouri. Clusters of this type contain firms that are members of the same extended product or value chain, including both backward linkages (inputs purchased from other industries) and forward linkages (outputs sold to other industries). Although there are many methods for determining clusters, this analysis will employ a statistical cluster analysis of input-output transactions within Missouri.

By taking the backward-linkage and forward-linkage clusters, the information technology value-chain can be ascertained. This information is significant in that decision-makers need to know where potential IT suppliers and consumers are located within the state. This allows businesses to better select facility locations, in that it identifies areas where IT suppliers and consumers are located. Also, it allows government officials to develop a strategy for recruiting IT firms, highlighting the state's existing supplier and consumer base. In general, the most optimal areas for the development of the IT industry are characterized by a sizable industry cluster employment base and high specialization. This information can be ascertained by determining the IT value-chain, and then analyzing the employment base and specialization within each cluster of the value-chain.

Industry Profile

Both nationally and in Missouri, the information technology (IT) industry has grown markedly over the last decade. IT employment in Missouri was estimated at 29,482 in 2000, an increase of 172.6% since 1990. Total annual wages during 2000 in the state's IT sector was estimated at $1,689.3 million, an increase of 264.5% since 1990. Estimated annual average wages per job during 2000 in the IT sector was $57,292, an increase of 33.7% since 1990. The number of Missouri IT firms in 2000 was estimated at 2,449, an increase of 171.0% since 1990.

In 2000, the IT industry accounted for 1.1% of total employment and 2.1% of total wages in Missouri. The annual average wage per job during 2000 in the IT sector was $57,292, well above the state average wage per job of $30,721. Refer to Table 1.

Table 1
Information Technology Economic Indicators, 1990-2000

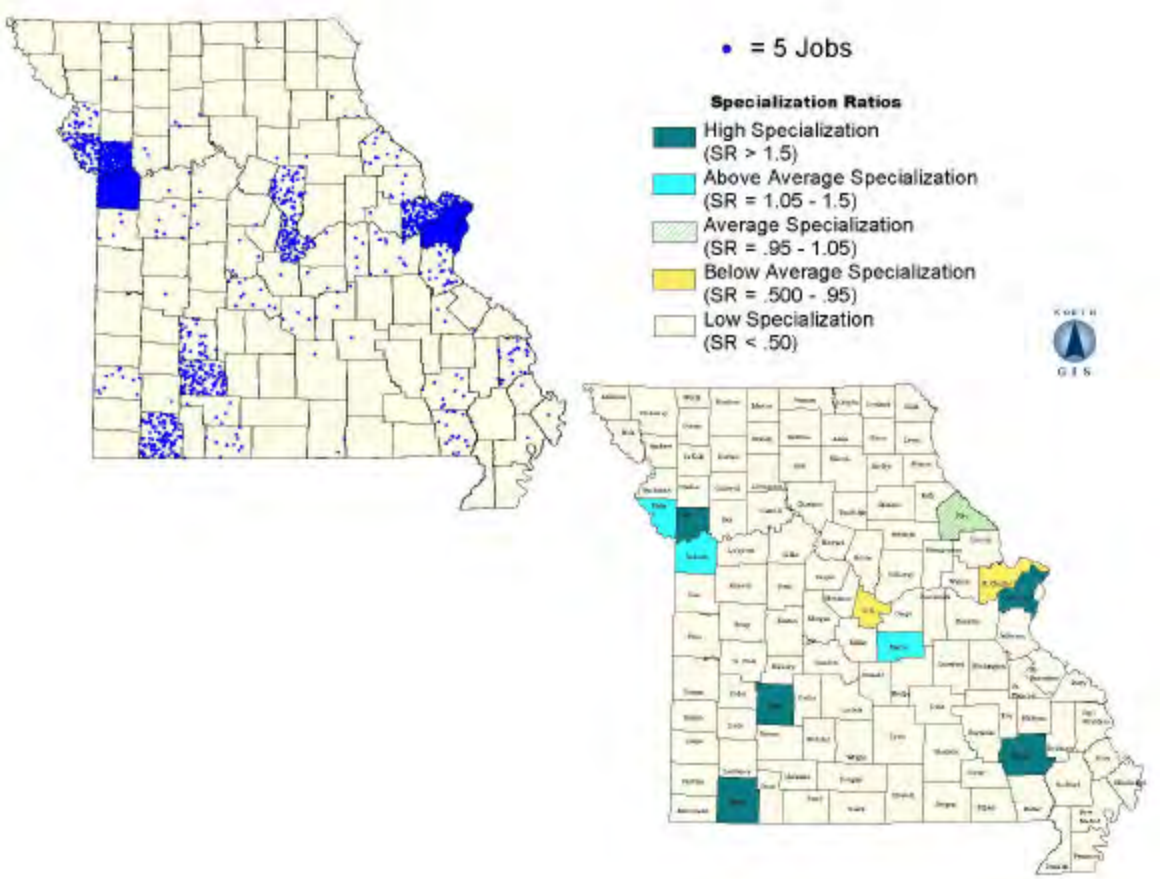
INDICATOR	1990	2000	Percent Change
Total Employment (Percent of Total Missouri Employment)	10,817 (0.5%)	29,482 (1.1%)	172.6% ↑
Total Wages, in Millions of 2000 Dollars (Percent of Total Missouri Wages)	$463.5 (0.7%)	$1,689.3 (2.1%)	264.5% ↑
Annual Average Wage Per Job, in 2000 Dollars (Missouri Annual Average Wage Per Job)	$42,862 ($28,462)	$57,292 ($30,721)	33.7% ↑
Total Firms	904	2,449	171.0% ↑

Source: Covered Employment and Wages, Missouri Department of Economic Development.

In 2000, the majority of IT jobs were located in metropolitan St. Louis, Kansas City, Columbia-Jefferson City, and Springfield. Additionally, IT jobs are also located in several rural areas of the state, such as Barry County, Polk County and Wayne County. Counties with the largest employment base were St. Louis (14,718), Jackson (6,586), Clay (2,046), St. Louis City (1,300), St. Charles (989) and Barry (635).

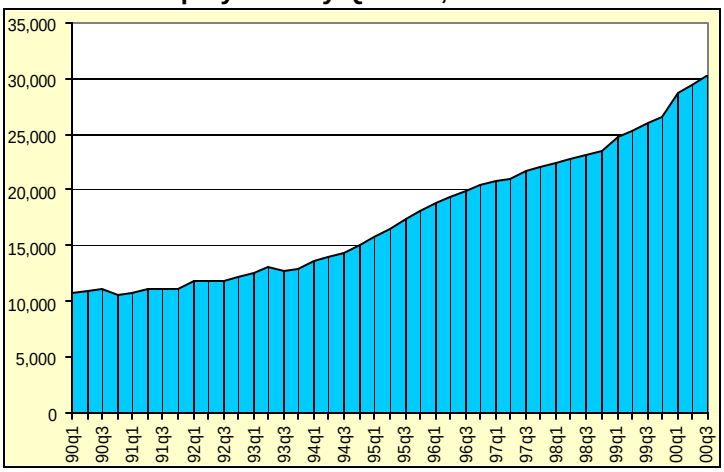
According to specialization ratios, 5 Missouri counties were highly specialized in IT employment. These areas were located in suburban metropolitan areas and in several rural areas of the state. The most specialized counties in the state were Barry (3.73), Clay (2.16), St. Louis (2.04), Polk (1.94) and Wayne (1.59). It is important to note that specialization ratios measure the proportion of industry employment relative to the state average, and not the total number of jobs. Refer to Map 1.

Map 1
Information Technology Employment and Specialization, 2000



Source: Covered Employment and Wages, Missouri Department of Economic Development.

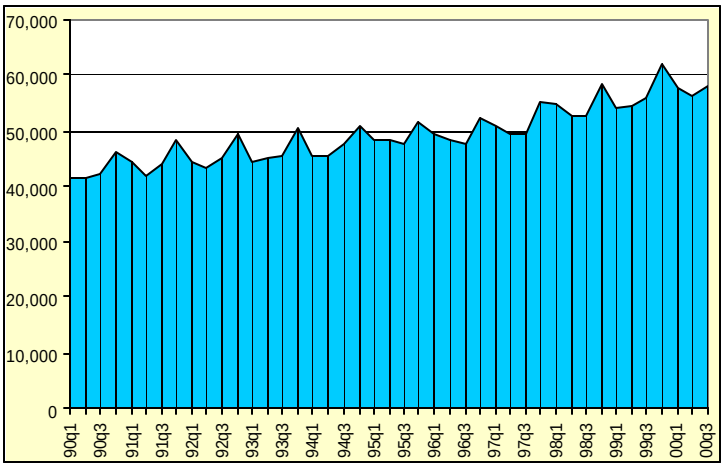
Employment by Quarter, 1990-2000



Source: Covered Employment and Wages, Missouri Dept of Economic Development.

- IT employment in Missouri has increased markedly during the last decade, with significant growth since 1994.
- The most current data estimates 30,349 IT jobs during 3rd quarter 2000.
- Employment was highest during 3rd quarter 2000, with 30,349 jobs.
- Employment was lowest during 4th quarter 1990, with 10,658 jobs.

Annual Average Wages Per Job by Quarter, 1990-2000

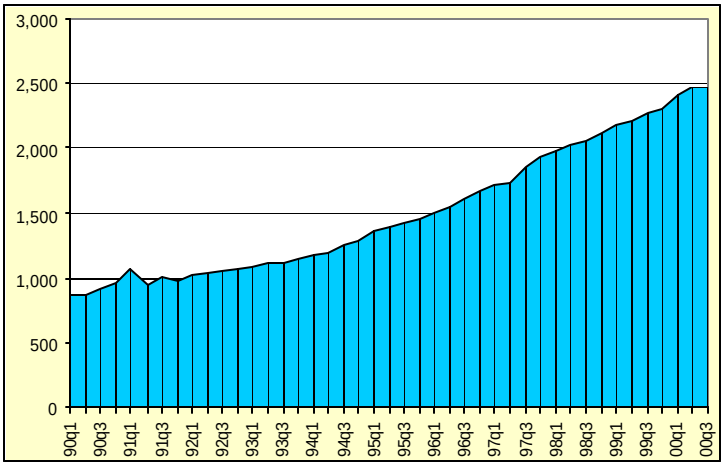


Source: Covered Employment and Wages, Missouri Dept of Economic Development.

- Annual average wages per job in the IT industry in Missouri has steadily increased since 1990.
- The most current data estimates an average wage per job of $58,082 during 3rd quarter 2000 - significantly higher than the state average wage of $30,497.
- Average wages per job were highest during 4th quarter 1999, at $62,172 per job.
- Average wages per job were lowest during 2nd quarter 1990, at $41,581 per job.

Adjusted to 2000 Real Dollars.
Quarter Wages Annualized.

Number of Firms by Quarter, 1990-2000



Source: Covered Employment and Wages, Missouri Dept of Economic Development.

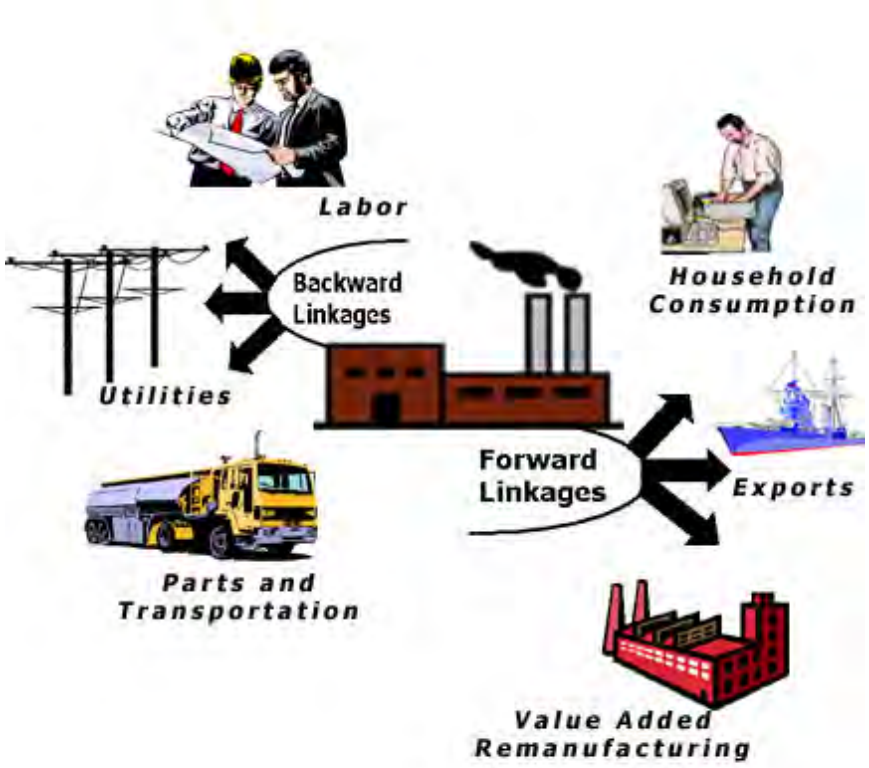
- The number of IT firms in Missouri has substantially increased since 1990 - especially since 1997.
- The most current data estimates 2,468 IT firms during 3rd quarter 2000.
- Firms were most numerous during 2000, peaking in 3rd quarter 2000 at 2,468 establishments.
- Firms were least numerous during 1st quarter 1990, at 870 establishments.

II. Methods

As stated previously, this analysis determines the value-chain clusters for the information technology (IT) industry in Missouri. Clusters of this type contain firms that are members of the same extended product or value chain, including both backward linkages (inputs purchased from other industries) and forward linkages (outputs sold to other industries). Refer to Figure 1.

The information technology industry is generally defined as firms engaged in providing IT services and software. Based on a typology developed by Stough (1998), IT is comprised of the following sectors: computer programming services (SIC 7371); prepackaged software (SIC 7372); computer integrated systems design (SIC 7373); computer and data processing (SIC 7374); information retrieval services (SIC 7375); computer facilities management (SIC 7376); computer rental and leasing (SIC 7377); computer maintenance and repair (SIC 7378); and other computer related services (SIC 7379).

Figure 1
Industry Value-Chain: Backward-Linkages and Forward-Linkages



Information technology industry clusters were determined by employing a hierarchical agglomerative statistical cluster analysis, using input-output trade flows among 525 industries in Missouri. This methodology for determining industry clusters has been used successfully by other research economists (Bergman and Feser 1999; Feser 1998). The most recent input-output (I-O) data is used, and is taken from the Minnesota IMPLAN Group. IMPLAN bases its data on the Bureau of Economic Analysis Benchmark Input-Output Study 1998. I-O data is the only major source of information on the interdependence of various industries in the United States. However, I-O data suffers from several drawbacks in that it is dated (most current data is from 1998), industry definitions may be imperfect, and it neglects supporting institutions. However, it is the only source of data on inter-industry transactions.

Cluster analysis is the generic name for a wide variety of procedures that can be used to create a classification. These procedures mathematically form clusters or groups of highly similar entities. More specifically, a clustering method is a multivariate statistical procedure that starts with data containing information about a sample of entities and attempts to reorganize these entities into relatively homogenous groups. In this case, the entities are purchases from and sales to 525 industries in Missouri - a 525x525 matrix.

In this analysis, a centroid hierarchical agglomerative cluster method was used to determine clusters. The centroid method essentially computes the center point of a polygon using existing cases as the boundaries of that polygon. The similarity of the cluster center point and a case under consideration is compared using some distance measure and, subsequently, joins the case to that cluster if a given level of similarity is achieved. The centroid method has a tendency to find relatively compact and hyperspherical clusters composed of highly similar cases (Aldenderfer and Blashfield 1984).

To compute the distance between centroids and cases, the Chebychev distance measure is used in this analysis. This dissimilarity measure is appropriate for continuous interval-ratio data. The Chebychev measure states that the distance between two cases is the maximum absolute difference between the values for the cases. Technically, all distance measures are best described as dissimilarity measures - a similarity measure (like a correlation coefficient) in reverse scale. Two points are identical if the distance between them is zero (Aldenderfer and Blashfield 1984).

Lastly, the cluster solutions obtained using the above mentioned cluster method and distance measure was compared to other solutions using alternative methods and measures: (1) centroid method using squared Euclidean distance; (2) Ward's method using squared Euclidean distance; and (3) Ward's method using Chebychev distance. All three methods yielded highly similar cluster solutions, indicating that there is an inherent structure in the data. Refer to Appendix A for more information.

III. Determination of Industry Clusters

Backward-Linkage Clusters

Backward-linkage clusters are groups of industries from which the information technology sector purchases goods and services. In essence, these industries supply the goods and services (inputs) needed for the IT industry to operate. Input trade flows among 525 industries in Missouri are analyzed to determine the backward-linkage clusters in the IT industry. This data measures purchase transactions between industries within Missouri.

The IT industry mainly purchases inputs from within its own industry - $240.6 million in purchases from computer and data processing. Other supply industries include wholesale trade ($48.6 million), electronic components ($41.1 million), and communications services ($40.7 million). Refer to Table 2.

Table 2
Information Technology Input Transactions, 1998
All Transactions Occur Within Missouri.

INDUSTRY PURCHASED FROM	INPUTS (Dollars)
Computer and Data Processing Services	240,574,400
Wholesale Trade	48,637,310
Electronic Components	41,115,090
Communications - Except Radio and TV	40,655,430
Real Estate	34,838,930
Other Business Services	25,301,870
Personnel Supply Services	17,188,740
Legal Services	16,819,010
Banking	11,479,550
U.S. Postal Service	11,300,320
Semiconductors and Related Devices	8,626,368
Management and Consulting Services	7,170,147
Commercial Printing	7,100,301
Accounting- Auditing and Bookkeeping	6,106,013
Hotels and Lodging Places	5,580,916
Colleges and Universities - Schools	5,430,478
Maintenance and Repair Other Facilities	5,324,950
Computer Storage Devices	5,206,694
Eating and Drinking	5,154,294
Security and Commodity Brokers	4,958,426

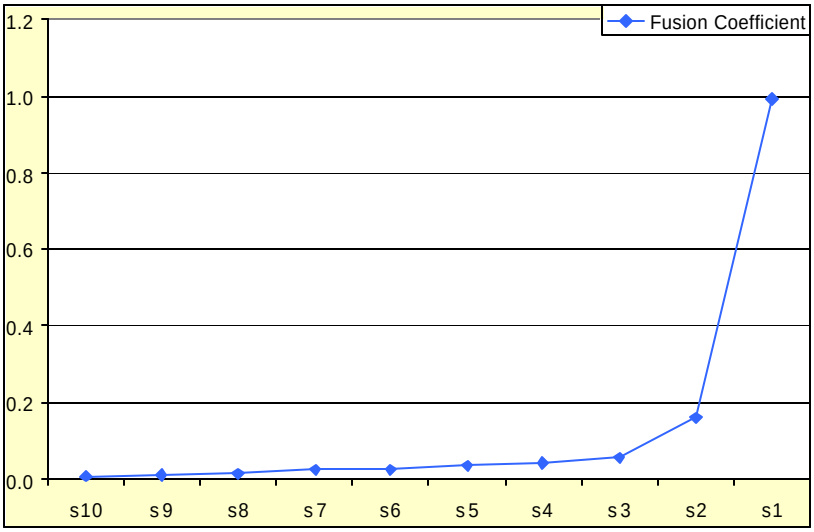
Source: Minnesota IMPLAN Group, based on Bureau of Economic Analysis Benchmark Input-Output.

To determine backward-linkage clusters, a hierarchical agglomerative cluster analysis was conducted using the 525x525 input-output matrix. The centroid method was used to form clusters since it tends to find relatively compact and hyperspherical clusters composed of highly similar cases. The Chebychev distance measure was used to compute dissimilarity between cases, since it is an appropriate distance measure for continuous interval-ratio data. Since the Chebychev distance measure is sensitive to extreme differences in scale, the variables have been standardized on a zero to one scale (although there is disagreement on this point, the technique is supported by Aldenderfer and Blashfield 1984).

The results of the clustering method indicate the presence of three clusters, as evidenced by fusion coefficients and the dendogram. Fusion coefficients are an index of the loss of information incurred when merging two clusters. There is a significant loss of information at the 2-stage cluster (Fusion=0.158, $Fusion_{change}$=0.1013), and convention dictates that the prior cluster stage indicates a good cluster solution (Aldenderfer and Blashfield 1984). Refer to Table 3 and Chart 1.

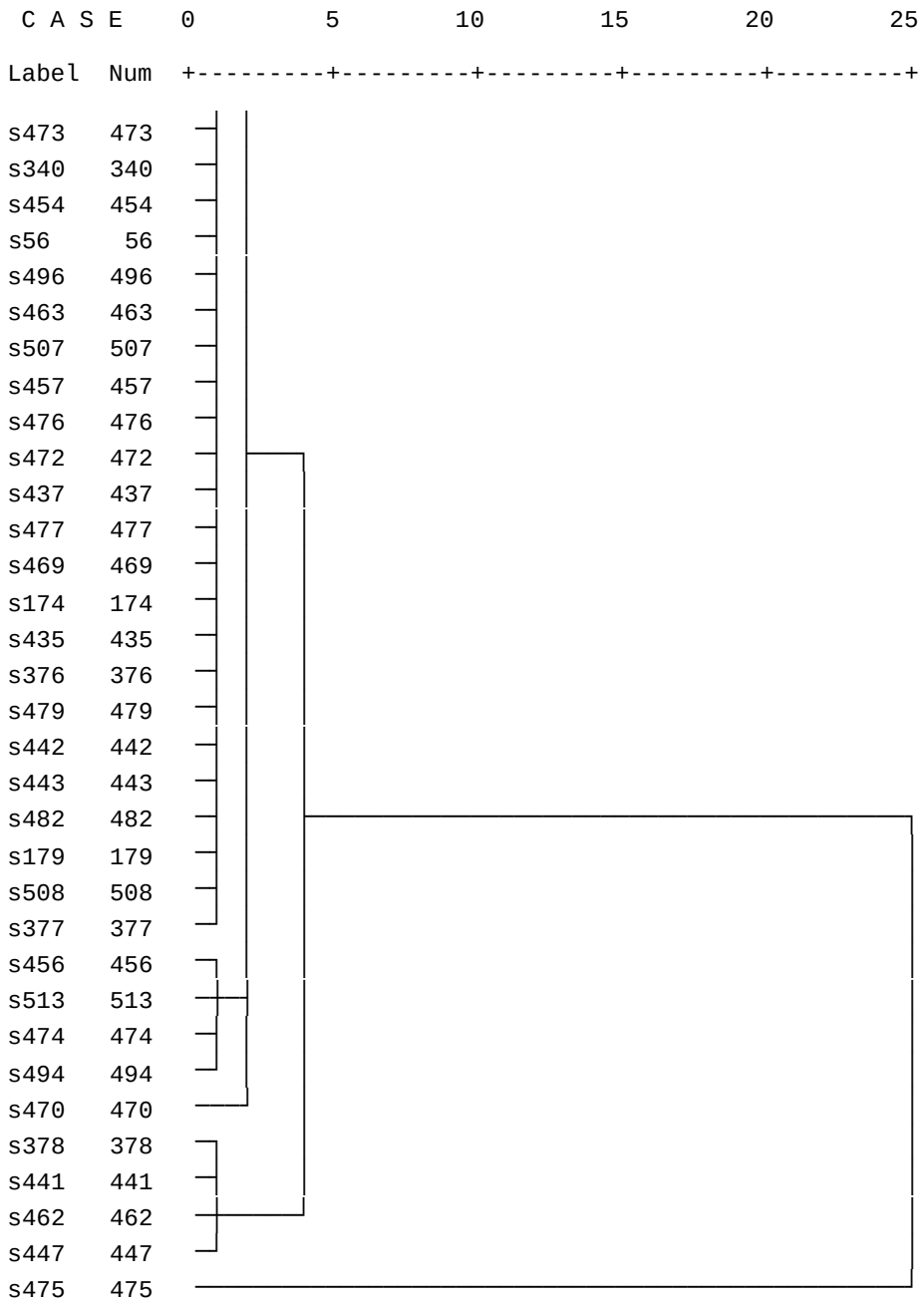
Table 3 and Chart 1
Fusion Coefficients, Agglomeration Schedule
Information Technology Backward-Linkage Clusters
Centroid Method Using Chebychev Distance.

CLUSTER STAGE	FUSION COEFFICIENT	LOSS OF INFORMATION
10	0.00612	0.0005
9	0.01053	0.0044
8	0.01483	0.0043
7	0.02277	0.0079
6	0.02466	0.0019
5	0.03480	0.0101
4	0.04018	0.0054
3	0.05498	0.0148
2	0.15800	0.1030
1	0.99400	0.8360



The dendogram also indicates a three cluster solution. Although dendograms are mainly heuristic devices, it provides an important validation of the cluster solution. The dendogram is presented in Figure 2.

Figure 2
Dendogram, Information Technology Backward-Linkage Clusters
Centroid Method Using Chebychev Distance.



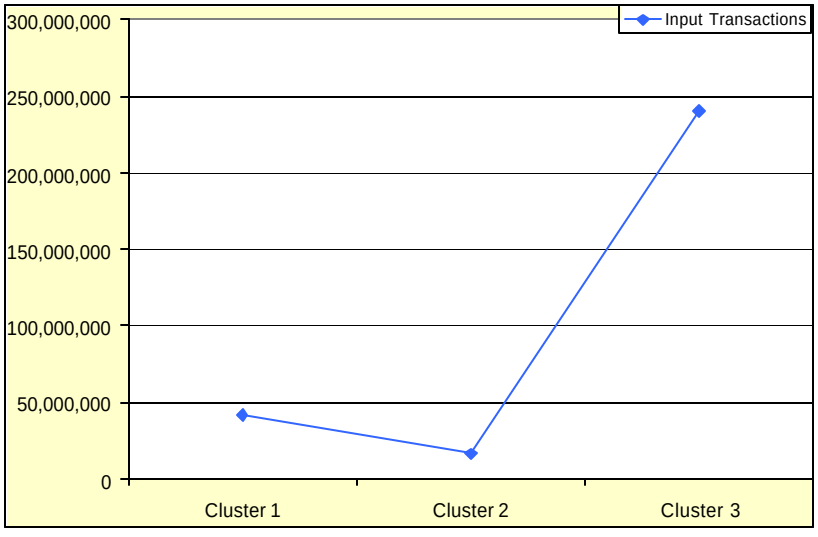
The analysis of variance (ANOVA) F-test indicates that the input variable is significantly different between the three clusters - statistically validating the cluster solution. By examining cluster membership and the differences in means, the three clusters of information technology inputs can be described. Refer to Table 4 and Chart 2.

The first cluster is described as supplying *Second-Order Inputs*, since the information technology industry purchases a moderate amount of goods and services from these industries (average purchases of $41.3 million). The second cluster is described as supplying *Third-Order Inputs*, since the information technology industry purchases a small amount of goods and services from these industries (average purchases of $16.4 million). The third cluster is described as supplying *First-Order Inputs*, since the information technology industry purchases a large amount of goods and services from these industries (average purchases of $240.6 million).

Table 4
Analysis of Variance, Information Technology Backward-Linkage Clusters

F-TEST			
F(3,521) = 14844.71 p = 0.000			
CLUSTER	N	MEAN	STANDARD DEVIATION
1	4	41,311,690.00	5,657,757.48
2	5	16,417,898.00	5,706,507.95
3	1	240,574,400.00	-

Chart 2
Cluster Means, Information Technology Backward-Linkage Clusters



The *First-Order Inputs* cluster is composed of industries from which the IT sector purchases the vast majority of its needed inputs (goods and services). It appears that the IT sector purchases inputs heavily from computer and data processing services. This cluster is composed of industries that support mission-critical operations of the IT industry.

The *Second-Order Inputs* cluster is composed of industries from which the IT sector purchases a moderate amount of its needed inputs (goods and services). It appears that the IT sector purchases inputs moderately from wholesale trade, electronic components manufacturing, communications and real estate. This cluster is composed of industries that support the physical operation and communication of the IT industry.

The *Third-Order Inputs* cluster is composed of industries from which the IT sector purchases a small amount of its needed inputs (goods and services). It appears that the IT sector purchases inputs sparingly from business services, personnel supply services, legal services, banking and the U.S. Postal Service. This cluster is composed of industries that support the financial and human resource components of the IT industry.

Table 5
Information Technology Backward-Linkage Clusters
All Transactions Occur Within Missouri.

CLUSTER	INDUSTRY	INPUTS (Dollars)
First-Order Inputs	Computer and Data Processing Services	240,574,400.00
Second-Order Inputs	Wholesale Trade	48,637,310.00
	Electronic Components	41,115,090.00
	Communications - Except Radio and TV	40,655,430.00
	Real Estate	34,838,930.00
Third-Order Inputs	Other Business Services	25,301,870.00
	Personnel Supply Services	17,188,740.00
	Legal Services	16,819,010.00
	Banking	11,479,550.00
	U.S. Postal Service	11,300,320.00

Forward-Linkage Clusters

Forward-linkage clusters are groups of industries to which the information technology sector sells goods and services. In essence, these industries purchase the goods and services produced by the IT industry (outputs). Output trade flows among 525 industries in Missouri are analyzed to determine the forward-linkage clusters in the IT industry. This data measures sales transactions between industries within Missouri.

The IT industry mainly sells outputs to wholesalers and other IT firms - $332.4 million in sales to wholesale trade, and $240.6 million in sales to computer and data processing services. The IT industry also sells to a host of other professional services firms: banks ($161.6 million), doctors and dentists ($138.1 million), communication services ($123.6 million), and colleges and universities ($100.9 million). Refer to Table 6.

Table 6
Information Technology Output Transactions, 1998
All Transactions Occur Within Missouri.

INDUSTRY PURCHASED FROM	OUTPUTS (Dollars)
Wholesale Trade	332,367,700
Computer and Data Processing Services	240,574,400
Banking	161,607,500
Doctors and Dentists	138,126,600
Communications - Except Radio and TV	123,618,200
Colleges and Universities - Schools	100,927,800
Hospitals	81,500,140
Engineering - Architectural Services	60,137,580
Air Transportation	54,178,270
Equipment Rental and Leasing	48,638,110
Other Business Services	46,014,230
Aircraft Manufacturing	44,587,310
Management and Consulting Services	44,241,520
Motor Freight Transport and Warehousing	36,750,770
Arrangement Of Passenger Transportation	34,991,800
Electric Services	28,137,990
Accounting - Auditing and Bookkeeping	26,351,470
Other Medical and Health Services	26,089,940
Cyclic Crudes - Industrial Organic Chemicals	23,978,180
Railroads and Related Services	23,270,670

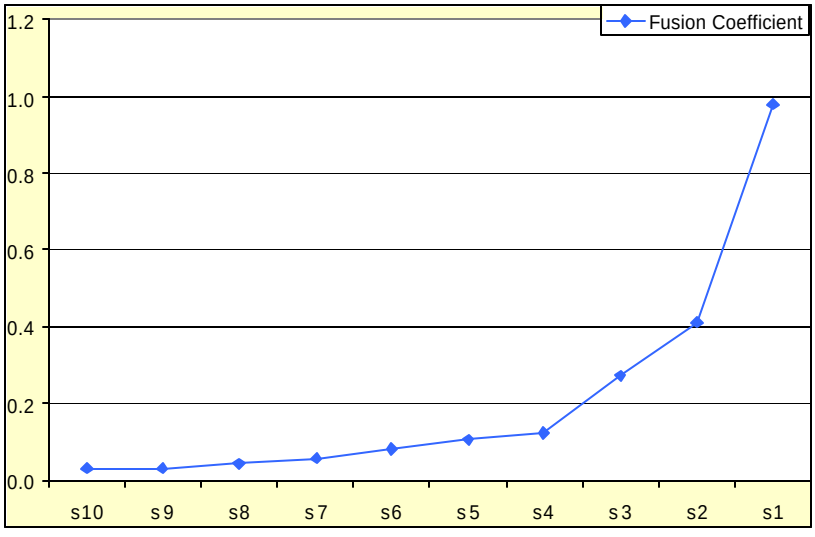
Source: Minnesota IMPLAN Group, based on Bureau of Economic Analysis Benchmark Input-Output.

To determine forward-linkage clusters, a hierarchical agglomerative cluster analysis was conducted using the 525x525 input-output matrix. The centroid method was used to form clusters since it tends to find relatively compact and hyperspherical clusters composed of highly similar cases. The Chebychev distance measure was used to compute dissimilarity between cases, since it is an appropriate distance measure for continuous interval-ratio data. Since the Chebychev distance measure is sensitive to extreme differences in scale, the variables have been standardized on a zero to one scale (although there is disagreement on this point, the technique is supported by Aldenderfer and Blashfield 1984).

The results of the clustering method indicate the presence of four clusters, as evidenced by fusion coefficients and the dendogram. Fusion coefficients are an index of the loss of information incurred when merging two clusters. There is a significant loss of information at the 3-stage cluster (Fusion=0.274, $Fusion_{change}$=0.148), and convention dictates that the prior cluster stage indicates a good cluster solution (Aldenderfer and Blashfield 1984). Refer to Table 7 and Chart 3.

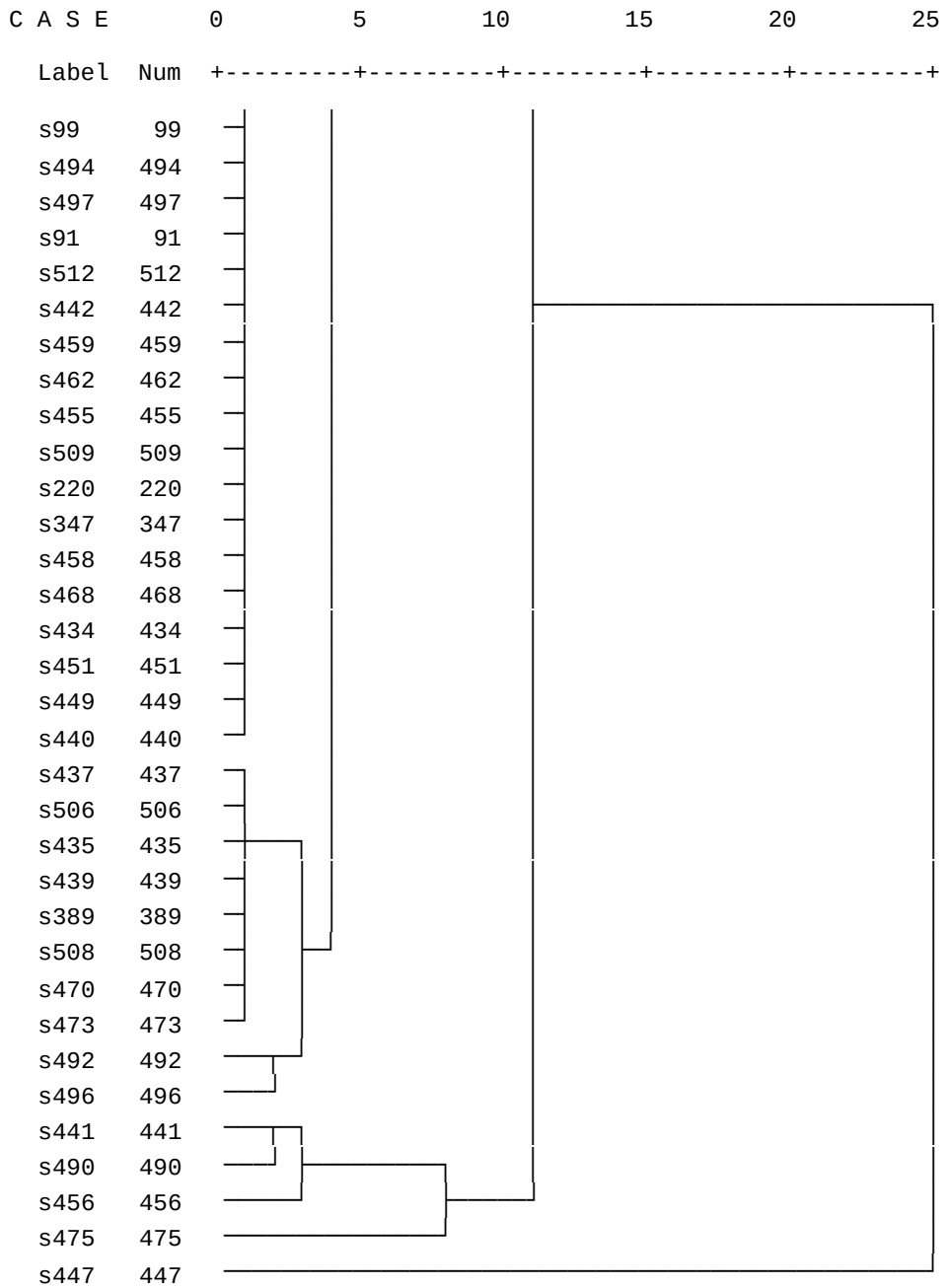
Table 7 and Chart 3
Fusion Coefficients, Agglomeration Schedule
Information Technology Forward-Linkage Clusters
Centroid Method Using Chebychev Distance.

CLUSTER STAGE	FUSION COEFFICIENT	LOSS OF INFORMATION
10	0.03145	0.0041
9	0.03218	0.0007
8	0.04365	0.0115
7	0.05845	0.0148
6	0.08156	0.0231
5	0.10800	0.0264
4	0.12600	0.0180
3	0.27400	0.1480
2	0.41300	0.1390
1	0.97700	0.5640



The dendogram also indicates a four cluster solution. Although dendograms are mainly heuristic devices, it provides an important validation of the cluster solution. The dendogram is presented in Figure 3.

Figure 3
Dendogram, Information Technology Forward-Linkage Clusters
Centroid Method Using Chebychev Distance.



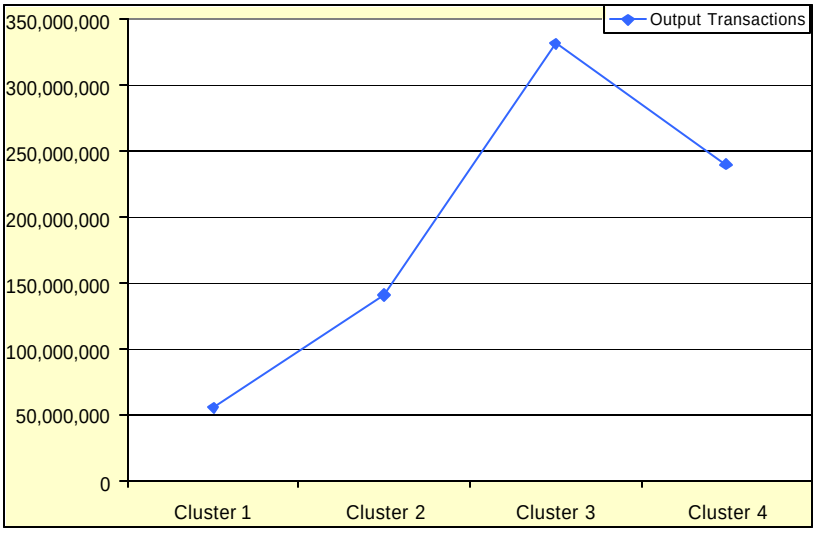
The analysis of variance (ANOVA) F-test indicates that the output variable is significantly different between the four clusters - statistically validating the cluster solution. By examining cluster membership and the differences in means, the four clusters of information technology outputs can be described. Refer to Table 8 and Chart 4.

The first cluster is described as supplying *Fourth-Order Outputs*, since the information technology industry sells a small amount of goods and services to these industries (average sales of $55.2 million). The second cluster is described as supplying *Third-Order Outputs*, since the information technology industry sells a moderate amount of goods and services to these industries (average sales of $141.1 million). The third cluster is described as supplying *First-Order Outputs*, since the information technology industry sells the most amount of goods and services to these industries (average sales of $332.4 million). The fourth cluster is described as supplying *Second-Order Outputs*, since the information technology industry sells a large amount of goods and services to these industries (average sales of $240.6 million).

Table 8
Analysis of Variance, Information Technology Forward-Linkage Clusters

F-TEST			
F(4,520) = 2528.14 p = 0.000			
CLUSTER	N	MEAN	STANDARD DEVIATION
1	10	55,196,753.00	20,862,119.46
2	3	141,117,433.33	19,170,434.04
3	1	332,367,700.00	-
4	1	240,574,400.00	-

Chart 4
Cluster Means, Information Technology Forward-Linkage Clusters



The *First-Order Outputs* cluster is composed of industries to which the IT sector sells the most of its outputs (goods and services). It appears that the IT sector sells heavily to the wholesale trade sector. This cluster is composed of wholesalers that purchase IT goods and services to sell to retailers.

The *Second-Order Outputs* cluster is composed of industries to which the IT sector sells a large share of its outputs (goods and services). It appears that the IT sector sells heavily to the computer and data processing industry. This cluster is composed of IT firms that purchase the goods and services of other IT firms to conduct business.

The *Third-Order Outputs* cluster is composed of industries to which the IT sector sells a moderate amount of its outputs (goods and services). It appears that the IT sector sells moderately to the banking sector, doctors and dentists, and the communications industry. This cluster is composed of professional services that require heavy investment in IT goods and services.

The *Fourth-Order Outputs* cluster is composed of industries to which the IT sector sells a small amount of its outputs (goods and services). It appears that the IT sector sells sparingly to higher education, hospitals, engineering and architectural firms, air transportation and manufacturing, and various business and consulting services. This cluster is composed of manufacturing and service industries that require moderate investment in IT goods and services.

Table 9
Information Technology Forward-Linkage Clusters
All Transactions Occur Within Missouri.

CLUSTER	INDUSTRY	OUTPUTS (Dollars)
First-Order Outputs	Wholesale Trade	332,367,700
Second-Order Outputs	Computer and Data Processing Services	240,574,400
Third-Order Outputs	Banking	161,607,500
	Doctors and Dentists	138,126,600
	Communications - Except Radio and TV	123,618,200
Fourth-Order Outputs	Colleges and Universities - Schools	100,927,800
	Hospitals	81,500,140
	Engineering and Architectural Services	60,137,580
	Air Transportation	54,178,270
	Equipment Rental and Leasing	48,638,110
	Other Business Services	46,014,230
	Aircraft Manufacturing	44,587,310
	Management and Consulting Services	44,241,520
	Motor Freight Transport and Warehousing	36,750,770
	Arrangement Of Passenger Transportation	34,991,800

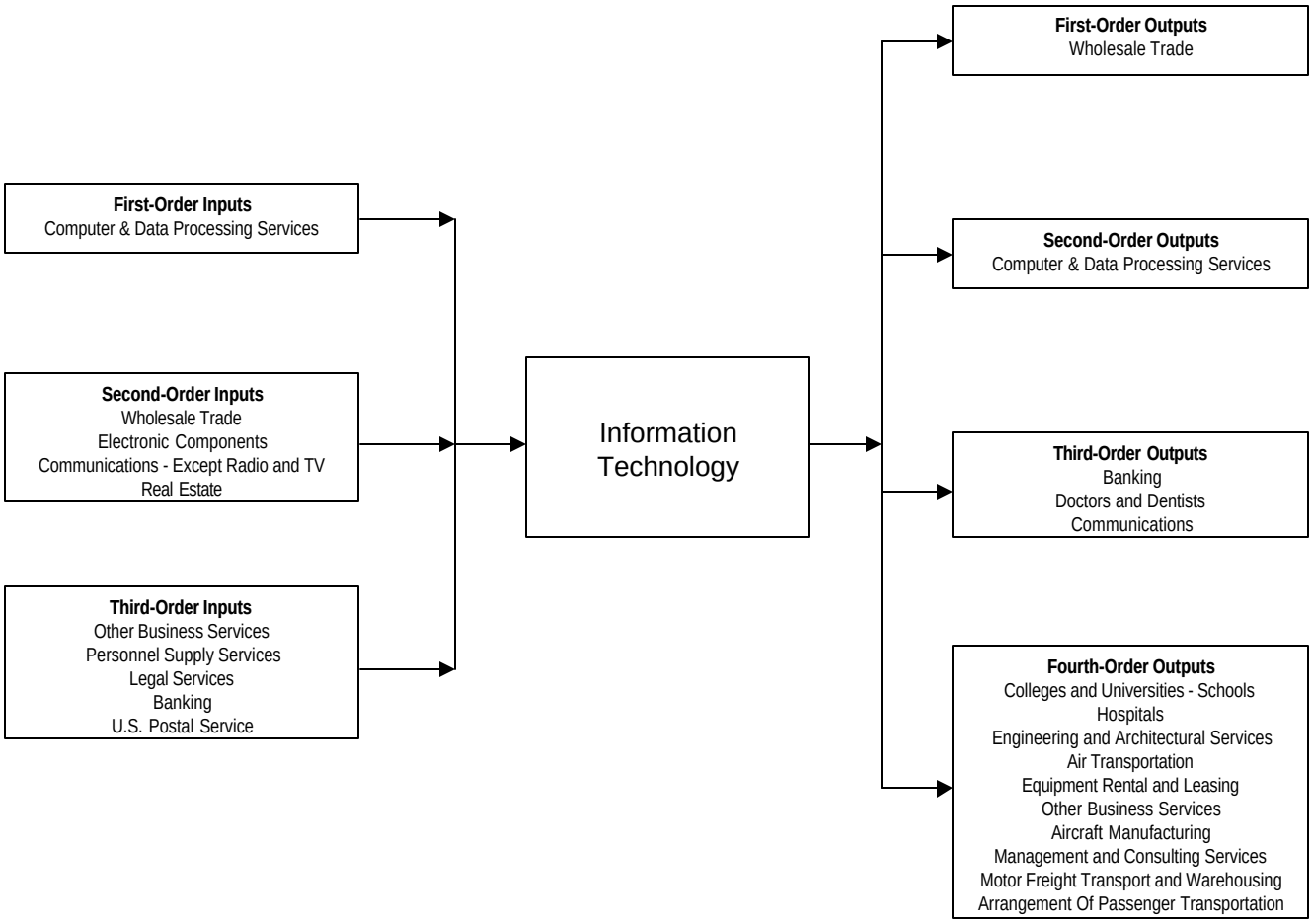
Industry Value-Chain

By taking the backward-linkage and forward-linkage clusters, the information technology value-chain can be ascertained. This is significant in that decision-makers need to know which industries are suppliers and consumers of IT goods and services. This will allow government officials to better identify and target which industries in Missouri supply the IT industry, and which industries purchase IT goods and services. This will permit state and local government to target programs aimed at developing these ancillary industries, especially supplier industries. Refer to Figure 4.

In terms of backward-linkages, it appears that the IT industry purchases most of its inputs from the computer and data processing sector - indicating that the IT sector is highly dependent on other IT firms to provide needed inputs. The IT industry also purchases moderately from: the wholesale trade and electronic components sectors to obtain various goods used in production; the communications sector to obtain telecommunications and internet services; and the real estate sector for facilities. Lastly, the IT industry purchases sparingly from a host of business operation services in support of personnel, fiscal, legal and shipping operations.

In terms of forward-linkages, it appears that the IT industry sells most of its outputs to the wholesale trade sector, who in turn sell IT goods to retailers for consumers to purchase. Additionally, the IT industry also sells heavily to other IT firms, indicating that there is a high degree of interdependence among IT firms. The IT industry also sells moderately to banks, the medical profession, and the telecommunications sector. Generally, the IT industry provides mission-critical services and software to these sectors that is specific to their business functions. Lastly, the IT industry sells sparingly to a host of professional service, transportation and manufacturing firms. The IT industry provides services and software to college and universities, hospitals, engineering and architectural firms, aircraft manufacturing and transportation firms, a variety of business and consulting services firms, and to the trucking and warehousing industry. Generally, these industries require a higher degree of IT infrastructure in order to operate.

Figure 4
Information Technology Industry Value-Chain



IV. Location of Industry Clusters

Now that we have delineated the information technology industry value-chain, we can begin to analyze where IT input and output clusters are located within the state. This is significant in that decision-makers need to know where potential IT suppliers and consumers are located in Missouri. This will allow businesses to better select facility locations, in that it identifies areas where suppliers and consumers are located. Also, it allows government officials to develop a strategy for recruiting IT firms, highlighting the region's existing supplier and consumer base.

In general, the most optimal areas for the IT industry to develop are characterized by a sizable industry cluster employment base and high specialization. This information can be ascertained by looking at the employment base and specialization within each cluster of the value-chain. Data is taken from Covered Employment and Wages, maintained by the Missouri Department of Economic Development.

Specialization ratios (SRs), also known as location quotients, are used to describe the dispersion of cluster employment across Missouri. SRs measure a county's employment concentration in a given cluster relative to the state average. Comparing these ratios over time gives an indication of the relative strengths and weaknesses of the industry cluster. SRs greater than 1.0 indicate that the county is relatively more specialized in an industry cluster relative to the state as a whole; or that the county has a comparative advantage in that cluster. SRs less than 1.0 indicate that the county is less specialized in an industry cluster relative to the state as a whole, which may indicate an area for potential growth; or that the county does not have a comparative advantage in that cluster.

It is important to note that SRs measure the proportion of industry cluster employment relative to the state average, and ***not*** the total number of jobs. Therefore, although St. Louis may have the largest number of employees within a particular industry cluster, it may account for only a small percentage of total employment – leading to a small SR. It is also important to note that the following SRs are normalized to the Missouri mean. In general, SRs are most informative when normalized to the national mean. However, national data was not available at this level of sector detail.

The formula for a SR is given below:

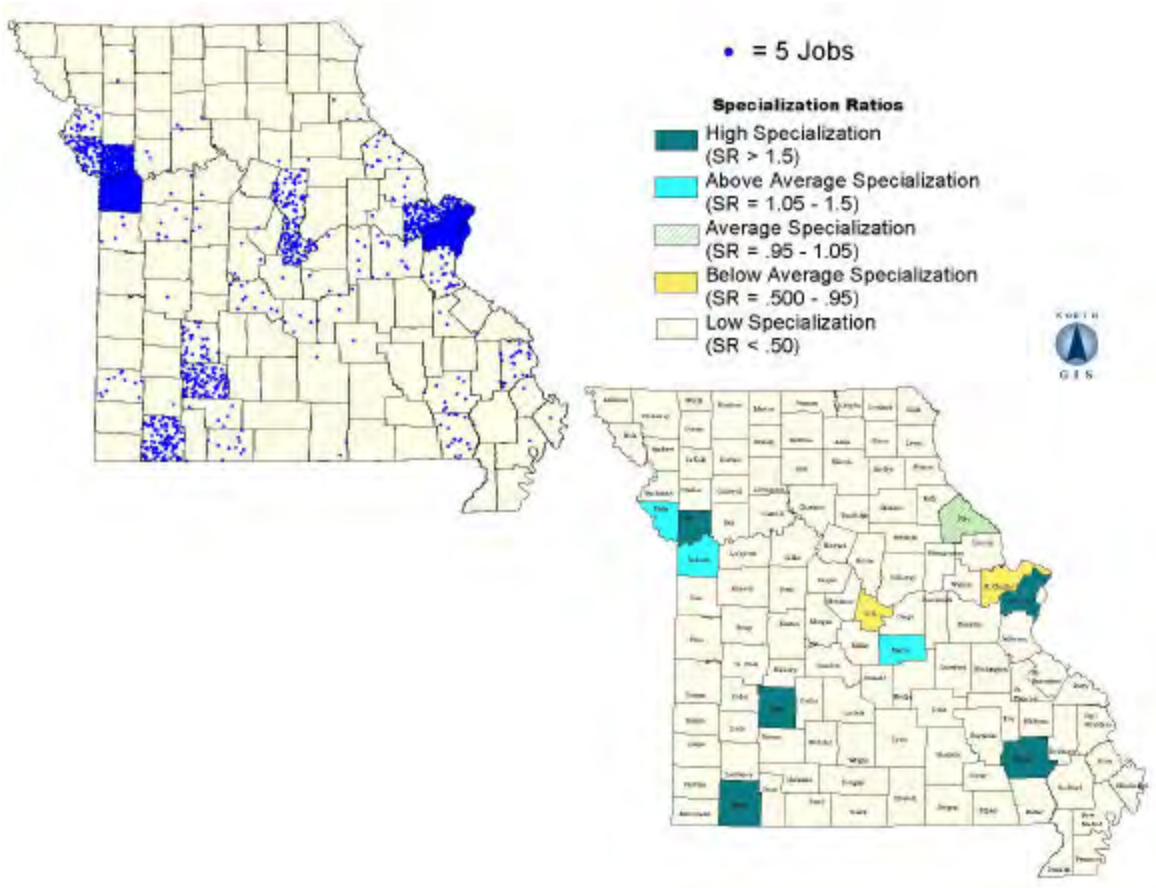
$$SR_{sector} = \left(\frac{\left(\frac{SECTOR\_EMPLOYMENT_{county}}{TOTAL\_EMPLOYMENT_{county}} \right)}{\left(\frac{SECTOR\_EMPLOYMENT_{state}}{TOTAL\_EMPLOYMENT_{state}} \right)} \right)$$

Backward-Linkage Clusters

In 2000, ***First-Order Input*** cluster jobs were located in metropolitan St. Louis, Kansas City, Columbia-Jefferson City, and Springfield. Additionally, jobs are also located in several rural areas of the state, including Barry County, Polk County and Wayne County. Counties with the largest employment base were St. Louis (14,718), Jackson (6,586), Clay (2,046), St. Louis City (1,300), and St. Charles (989).

According to specialization ratios, 5 Missouri counties were highly specialized in First-Order Input cluster employment. These areas were located in suburban metropolitan areas and in several rural areas of the state. The most specialized counties in the state were Barry (3.73), Clay (2.16), St. Louis (2.04), Polk (1.94) and Wayne (1.59). It is important to note that specialization ratios measure the proportion of cluster employment relative to the state average, and not the total number of jobs. Refer to Map 2 and Tables 10 and 11.

Map 2
First-Order Input Cluster - Employment and Specialization, 2000
Includes Computer & Data Processing Services.



Source: Covered Employment and Wages, Missouri Department of Economic Development.

Table 10
First-Order Input Cluster - Employment, 2000
Includes Computer & Data Processing Services.

COUNTY	EMPLOYMENT	SPECIALIZATION RATIO
St. Louis	14,717.56	2.04
Jackson	6,586.11	1.49
Clay	2,045.67	2.16
St. Louis City	1,300.33	0.46
St. Charles	989.44	0.93
Barry	634.78	3.73
Greene	601.56	0.38
Platte	476.00	1.24
Cole	432.44	0.75
Boone	370.00	0.44
Polk	155.11	1.94
Cape Girardeau	126.22	0.28
Jefferson	114.33	0.24
Buchanan	93.78	0.19
Butler	79.33	0.40

Source: Covered Employment and Wages, Missouri Department of Economic Development.

Table 11
First-Order Input Cluster - Specialization Ratios, 2000
Includes Computer & Data Processing Services.

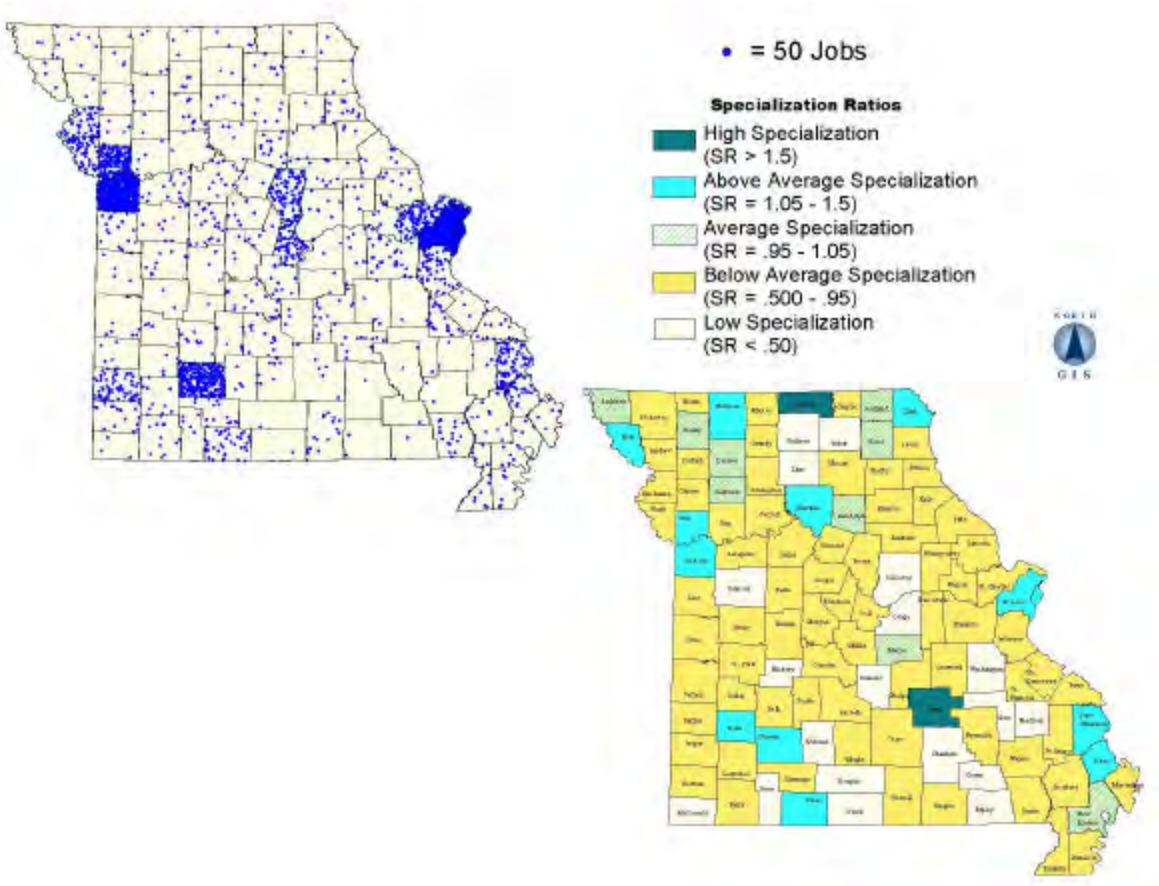
COUNTY	EMPLOYMENT	SPECIALIZATION RATIO
Barry	634.78	3.73
Clay	2,045.67	2.16
St. Louis	14,717.56	2.04
Polk	155.11	1.94
Wayne	44.00	1.59
Jackson	6,586.11	1.49
Platte	476.00	1.24
Maries	17.33	1.09
Pike	67.56	0.99
St. Charles	989.44	0.93
Cole	432.44	0.75
St. Louis City	1,300.33	0.46
Boone	370.00	0.44
Butler	79.33	0.40
Greene	601.56	0.38

Source: Covered Employment and Wages, Missouri Department of Economic Development.

In 2000, ***Second-Order Input*** cluster jobs were located in metropolitan St. Louis, Kansas City and Springfield. Additionally, jobs are also located in several rural areas of the state, including Cape Girardeau County, Cole County and Taney County. Counties with the largest employment base were St. Louis (62,907), Jackson (41,658), St. Louis City (24,044), Greene (14,358) and Clay (8,865).

According to specialization ratios, 2 Missouri counties were highly specialized in Second-Order Input cluster employment. These areas were located in two rural areas of the state. The most specialized counties in the state were Dent (1.82) and Putnam (1.54). It is important to note that specialization ratios measure the proportion of cluster employment relative to the state average, and not the total number of jobs. Refer to Map 3 and Tables 12 and 13.

Map 3
Second-Order Input Cluster - Employment and Specialization, 2000
Includes Wholesale Trade; Electronic Components; Communications; Real Estate.



Source: Covered Employment and Wages, Missouri Department of Economic Development.

Table 12
Second-Order Input Cluster - Employment, 2000
Includes Wholesale Trade; Electronic Components; Communications; Real Estate.

COUNTY	EMPLOYMENT	SPECIALIZATION RATIO
St. Louis	62,907.11	1.17
Jackson	41,658.00	1.27
St. Louis City	24,043.89	1.15
Greene	14,357.56	1.21
Clay	8,865.44	1.26
St. Charles	6,347.56	0.80
Boone	5,285.00	0.84
Cape Girardeau	4,193.67	1.26
Jasper	3,421.67	0.71
Cole	2,996.78	0.70
Platte	2,661.56	0.93
Buchanan	2,653.78	0.74
Jefferson	2,544.22	0.72
Taney	2,120.44	1.09
Franklin	1,601.33	0.55

Source: Covered Employment and Wages, Missouri Department of Economic Development.

Table 13
Second-Order Input Cluster - Specialization Ratios, 2000
Includes Wholesale Trade; Electronic Components; Communications; Real Estate.

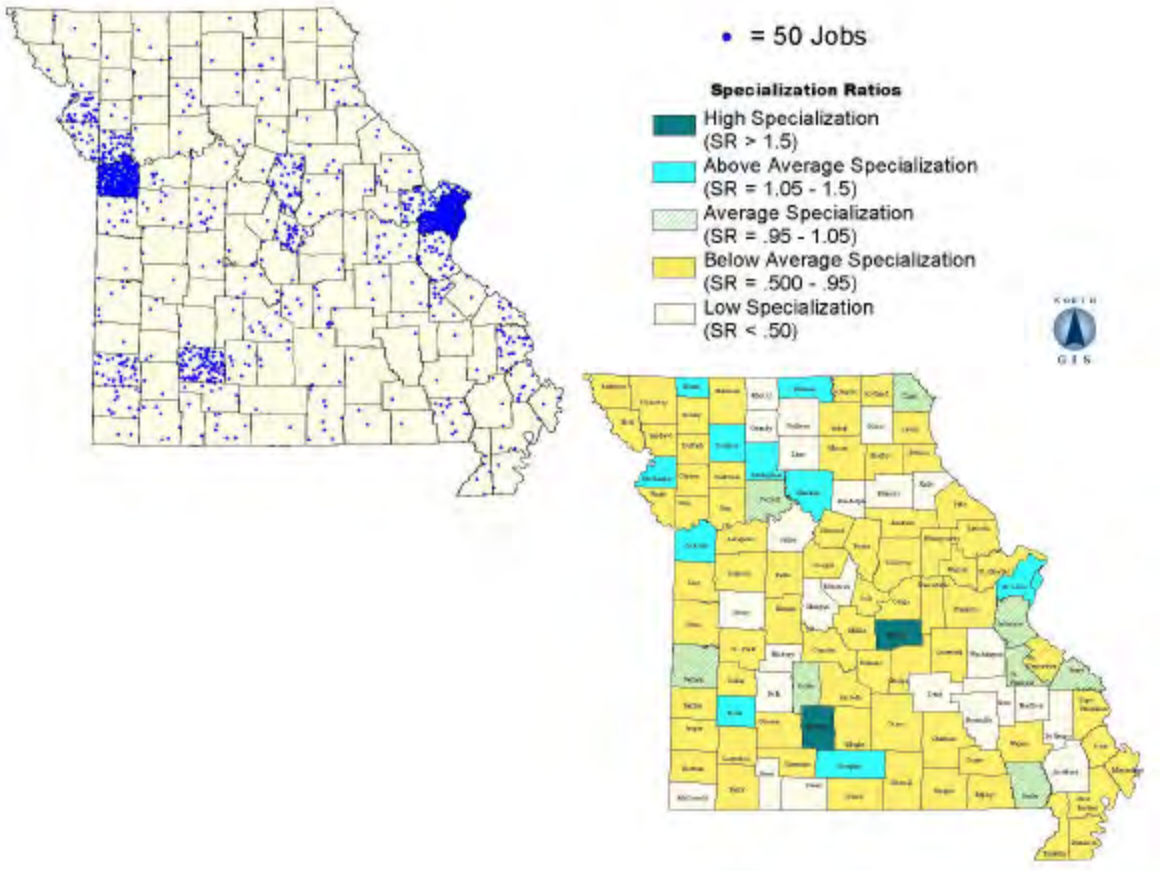
COUNTY	EMPLOYMENT	SPECIALIZATION RATIO
Dent	658.44	1.82
Putnam	132.22	1.54
Holt	161.67	1.37
Clark	171.89	1.35
Chariton	234.44	1.31
Jackson	41,658.00	1.27
Clay	8,865.44	1.26
Cape Girardeau	4,193.67	1.26
Greene	14,357.56	1.21
Scott	1,477.89	1.17
St. Louis	62,907.11	1.17
St. Louis City	24,043.89	1.15
Harrison	266.33	1.14
Taney	2,120.44	1.09
Dade	168.56	1.06

Source: Covered Employment and Wages, Missouri Department of Economic Development.

In 2000, ***Third-Order Input*** cluster jobs were located in metropolitan St. Louis, Kansas City, Springfield and Columbia. Additionally, jobs are also located in several rural areas of the state, including Cole County, Cape Girardeau County and St. Francois County. Counties with the largest employment base were St. Louis (27,343), Jackson (22,650), St. Louis City (12,639), Greene (4,959) and St. Charles (2,672).

According to specialization ratios, 2 Missouri counties were highly specialized in Third-Order Input cluster employment. These areas were located in the south central area of the state. The most specialized counties in the state were Webster (1.81) and Maries (1.66). It is important to note that specialization ratios measure the proportion of cluster employment relative to the state average, and not the total number of jobs. Refer to Map 4 and Tables 14 and 15.

Map 4
Third-Order Input Cluster - Employment and Specialization, 2000
Includes Other Business Services; Personnel Supply Services; Legal Services; Banking; US Postal Service.



Source: Covered Employment and Wages, Missouri Department of Economic Development.

Table 14
Third-Order Input Cluster - Employment, 2000
Includes Other Business Services; Personnel Supply Services; Legal Services; Banking; US Postal Service.

COUNTY	EMPLOYMENT	SPECIALIZATION RATIO
St. Louis	27,342.78	1.10
Jackson	22,650.22	1.49
St. Louis City	12,638.56	1.31
Greene	4,959.33	0.90
St. Charles	2,671.78	0.73
Boone	2,410.22	0.83
Clay	2,158.33	0.66
Buchanan	1,968.22	1.19
Jasper	1,866.44	0.83
Cole	1,681.44	0.84
Jefferson	1,625.78	0.99
Cape Girardeau	1,347.44	0.87
Franklin	909.44	0.68
St. Francois	749.78	0.99
Platte	680.00	0.51

Source: Covered Employment and Wages, Missouri Department of Economic Development.

Table 15
Third-Order Input Cluster - Specialization Ratios, 2000
Includes Other Business Services; Personnel Supply Services; Legal Services; Banking; US Postal Service.

COUNTY	EMPLOYMENT	SPECIALIZATION RATIO
Webster	445.00	1.81
Maries	91.00	1.66
Putnam	59.78	1.50
Jackson	22,650.22	1.49
St. Louis City	12,638.56	1.31
Chariton	103.78	1.25
Dade	90.33	1.23
Buchanan	1,968.22	1.19
Livingston	293.00	1.17
Daviess	82.67	1.14
St. Louis	27,342.78	1.10
Worth	18.44	1.08
Douglas	103.00	1.07
Clark	61.67	1.05
Perry	341.44	1.00

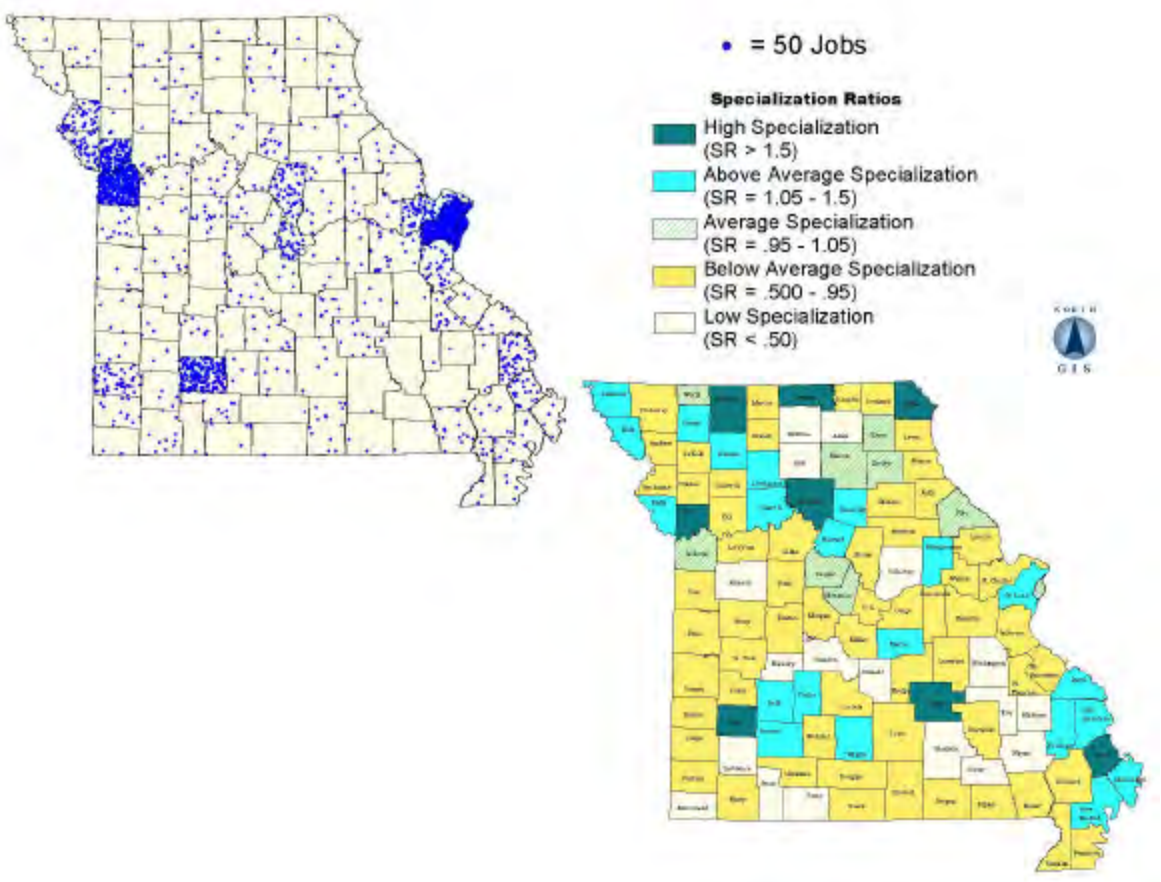
Source: Covered Employment and Wages, Missouri Department of Economic Development.

Forward-Linkage Clusters

In 2000, ***First-Order Output*** cluster jobs were located in metropolitan St. Louis, Kansas City and Springfield. Additionally, jobs are also located in several rural areas of the state, including Cape Girardeau County, Cole County and Scott County. Counties with the largest employment base were St. Louis (41,409), Jackson (21,452), St. Louis City (13,590), Greene (10,175) and Clay (7,706).

According to specialization ratios, 8 Missouri counties were highly specialized in First-Order Output cluster employment. These areas were located Kansas City and in several rural areas of the state. The most specialized counties in the state were Dent (2.68), Putnam (2.05), Clark (1.91), Chariton (1.77), Clay (1.70), Scott (1.53), Harrison (1.53) and Dade (1.51). It is important to note that specialization ratios measure the proportion of cluster employment relative to the state average, and not the total number of jobs. Refer to Map 5 and Tables 16 and 17.

Map 5
First-Order Output Cluster - Employment and Specialization, 2000
Includes Wholesale Trade.



Source: Covered Employment and Wages, Missouri Department of Economic Development.

Table 16
First-Order Output Cluster - Employment, 2000
Includes Wholesale Trade.

COUNTY	EMPLOYMENT	SPECIALIZATION RATIO
St. Louis	41,409.22	1.19
Jackson	21,452.22	1.02
St. Louis City	13,590.33	1.01
Greene	10,174.89	1.33
Clay	7,705.67	1.70
St. Charles	3,712.22	0.73
Boone	3,096.89	0.76
Jasper	2,601.22	0.84
Cape Girardeau	2,412.11	1.12
Platte	2,236.67	1.21
Buchanan	2,095.78	0.91
Cole	1,934.56	0.70
Jefferson	1,781.67	0.78
Scott	1,242.56	1.53
Franklin	1,170.56	0.63

Source: Covered Employment and Wages, Missouri Department of Economic Development.

Table 17
First-Order Output Cluster - Specialization Ratios, 2000
Includes Wholesale Trade.

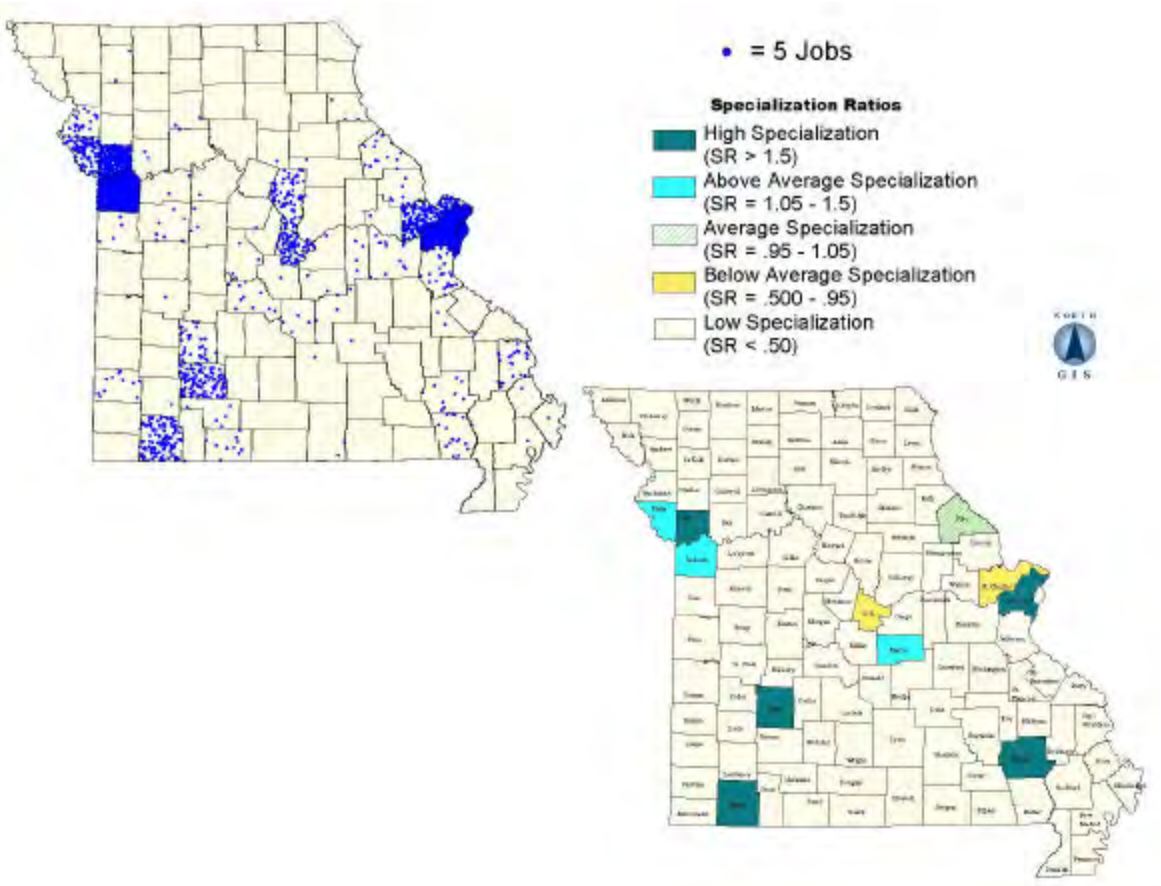
COUNTY	EMPLOYMENT	SPECIALIZATION RATIO
Dent	623.33	2.68
Putnam	113.44	2.05
Clark	156.33	1.91
Chariton	204.89	1.77
Clay	7,705.67	1.70
Scott	1,242.56	1.53
Harrison	230.33	1.53
Dade	154.33	1.51
Maries	109.89	1.44
New Madrid	614.89	1.38
Daviess	138.56	1.37
Greene	10,174.89	1.33
Montgomery	241.22	1.31
Holt	96.56	1.27
Bollinger	132.56	1.26

Source: Covered Employment and Wages, Missouri Department of Economic Development.

In 2000, ***Second-Order Output*** cluster jobs were located in metropolitan St. Louis, Kansas City, Columbia-Jefferson City, and Springfield. Additionally, jobs are also located in several rural areas of the state, including Barry County, Polk County and Wayne County. Counties with the largest employment base were St. Louis (14,718), Jackson (6,586), Clay (2,046), St. Louis City (1,300), St. Charles (989) and Barry (635).

According to specialization ratios, 5 Missouri counties were highly specialized in Second-Order Output cluster employment. These areas were located in the suburban metropolitan areas and in several rural areas of the state. The most specialized counties in the state were Barry (3.73), Clay (2.16), St. Louis (2.04), Polk (1.94) and Wayne (1.59). It is important to note that specialization ratios measure the proportion of cluster employment relative to the state average, and not the total number of jobs. Refer to Map 6 and Tables 18 and 19.

Map 6
Second-Order Output Cluster - Employment and Specialization, 2000
Includes Computer & Data Processing Services.



Source: Covered Employment and Wages, Missouri Department of Economic Development.

Table 18
Second-Order Output Cluster - Employment, 2000
Includes Computer & Data Processing Services.

COUNTY	EMPLOYMENT	SPECIALIZATION RATIO
St. Louis	14,717.56	2.04
Jackson	6,586.11	1.49
Clay	2,045.67	2.16
St. Louis City	1,300.33	0.46
St. Charles	989.44	0.93
Barry	634.78	3.73
Greene	601.56	0.38
Platte	476.00	1.24
Cole	432.44	0.75
Boone	370.00	0.44
Polk	155.11	1.94
Cape Girardeau	126.22	0.28
Jefferson	114.33	0.24
Buchanan	93.78	0.19
Butler	79.33	0.40

Source: Covered Employment and Wages, Missouri Department of Economic Development.

Table 19
Second-Order Output Cluster - Specialization Ratios, 2000
Includes Computer & Data Processing Services.

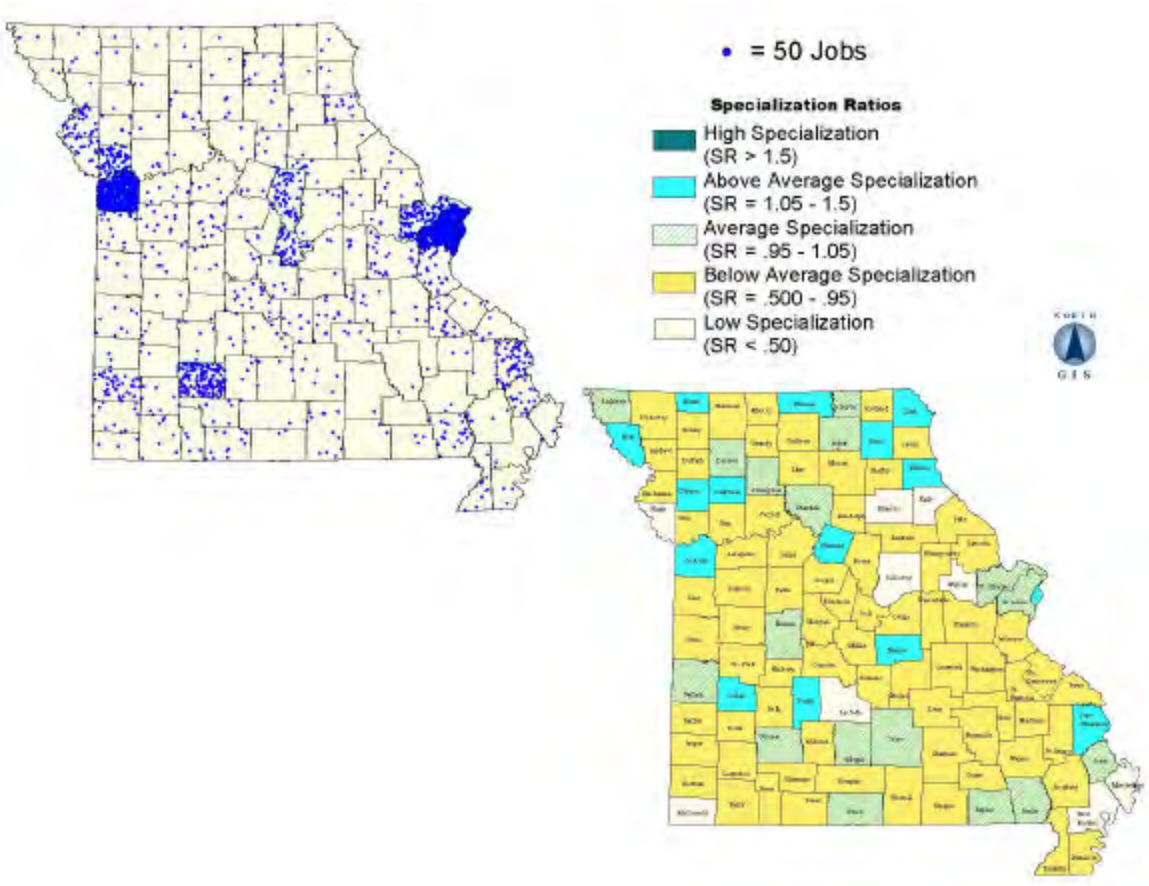
COUNTY	EMPLOYMENT	SPECIALIZATION RATIO
Barry	634.78	3.73
Clay	2,045.67	2.16
St. Louis	14,717.56	2.04
Polk	155.11	1.94
Wayne	44.00	1.59
Jackson	6,586.11	1.49
Platte	476.00	1.24
Maries	17.33	1.09
Pike	67.56	0.99
St. Charles	989.44	0.93
Cole	432.44	0.75
St. Louis City	1,300.33	0.46
Boone	370.00	0.44
Butler	79.33	0.40
Greene	601.56	0.38

Source: Covered Employment and Wages, Missouri Department of Economic Development.

In 2000, ***Third-Order Output*** cluster jobs were located in metropolitan St. Louis, Kansas City, Springfield and Columbia. Additionally, jobs are also located in several rural areas of the state, including Cape Girardeau County, Cole County and Butler County. Counties with the largest employment base were St. Louis (37,380), Jackson (32,875), St. Louis City (17,520), Greene (7,993) and St. Charles (5,376).

According to specialization ratios, no Missouri counties were highly specialized in Third-Order Output cluster employment. However, Maries County (1.49) and Jackson County (1.49) had significant above average specialization. It is important to note that specialization ratios measure the proportion of cluster employment relative to the state average, and not the total number of jobs. Refer to Map 7 and Tables 20 and 21.

Map 7
Third-Order Output Cluster - Employment and Specialization, 2000
Includes Banking; Doctors & Dentists; Communications.



Source: Covered Employment and Wages, Missouri Department of Economic Development.

Table 20
Third-Order Output Cluster - Employment, 2000
Includes Banking; Doctors & Dentists; Communications.

COUNTY	EMPLOYMENT	SPECIALIZATION RATIO
St. Louis	37,379.89	1.03
Jackson	32,874.56	1.49
St. Louis City	17,520.11	1.25
Greene	7,993.33	1.01
St. Charles	5,376.33	1.01
Clay	3,745.11	0.79
Boone	3,116.67	0.74
Cape Girardeau	3,018.89	1.35
Cole	2,594.89	0.90
Jasper	2,209.67	0.68
Buchanan	1,748.00	0.73
Jefferson	1,449.67	0.61
Franklin	1,437.78	0.74
Butler	1,016.78	1.03
Platte	947.89	0.49

Source: Covered Employment and Wages, Missouri Department of Economic Development.

Table 21
Third-Order Output Cluster - Specialization Ratios, 2000
Includes Banking; Doctors & Dentists; Communications.

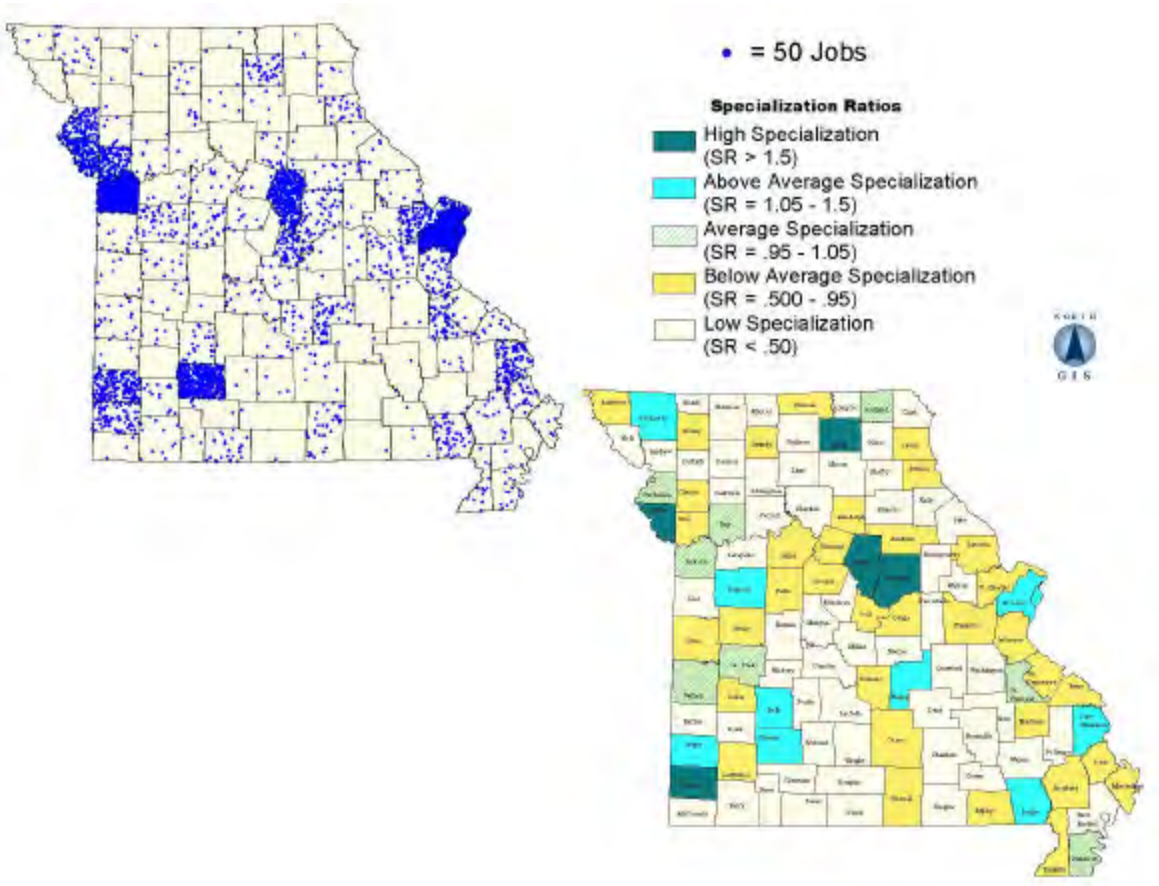
COUNTY	EMPLOYMENT	SPECIALIZATION RATIO
Maries	118.56	1.49
Jackson	32,874.56	1.49
Worth	34.56	1.40
Cape Girardeau	3,018.89	1.35
St. Louis City	17,520.11	1.25
Howard	166.33	1.25
Knox	70.78	1.22
Clark	101.00	1.19
Dallas	165.22	1.16
Cedar	193.89	1.14
Marion	892.33	1.14
Putnam	64.78	1.12
Clinton	283.22	1.11
Holt	87.00	1.10
Caldwell	87.33	1.09

Source: Covered Employment and Wages, Missouri Department of Economic Development.

In 2000, ***Fourth-Order Output*** cluster jobs were located in metropolitan St. Louis, Kansas City, Springfield and Columbia. Additionally, jobs are also located in several rural areas of the state, including Cape Girardeau County, Cole County and Callaway County. Counties with the largest employment base were St. Louis (83,445), Jackson (45,121), St. Louis City (40,573), Greene (23,509) and Boone (19,969).

According to specialization ratios, 5 Missouri counties were highly specialized in Fourth-Order Output cluster employment. These areas were located in the larger regional cities in Missouri. The most specialized counties in the state were Boone (2.21), Platte (2.15), Newton (1.86), Callaway (1.82), and Adair (1.57). It is important to note that specialization ratios measure the proportion of cluster employment relative to the state average, and not the total number of jobs. Refer to Map 8 and Tables 22 and 23.

Map 8
Fourth-Order Output Cluster - Employment and Specialization, 2000
Includes Colleges & Universities; Hospitals; Engineering & Architectural Services; Air Transportation; Equipment Rental & Leasing; Other Business Services; Aircraft Manufacturing; Management & Consulting Services; Motor Freight & Warehousing; Arrangement of Passenger Transportation.



Source: Covered Employment and Wages, Missouri Department of Economic Development.

Table 22
Fourth-Order Output Cluster - Employment, 2000
Includes Colleges & Universities; Hospitals; Engineering & Architectural Services; Air Transportation; Equipment Rental & Leasing; Other Business Services; Aircraft Manufacturing; Management & Consulting Services; Motor Freight & Warehousing; Arrangement of Passenger Transportation.

COUNTY	EMPLOYMENT	SPECIALIZATION RATIO
St. Louis	83,444.89	1.08
Jackson	45,121.22	0.96
St. Louis City	40,573.44	1.36
Greene	23,509.11	1.39
Boone	19,969.44	2.21
Jasper	9,075.78	1.31
Platte	8,794.11	2.15
Clay	6,748.22	0.67
St. Charles	5,826.56	0.51
Cape Girardeau	5,479.11	1.15
Buchanan	5,392.78	1.05
Newton	4,108.89	1.86
Cole	3,399.11	0.55
Callaway	2,960.44	1.82
Jefferson	2,774.89	0.55

Source: Covered Employment and Wages, Missouri Department of Economic Development.

Table 23
Fourth-Order Output Cluster - Specialization Ratios, 2000
Includes Colleges & Universities; Hospitals; Engineering & Architectural Services; Air Transportation; Equipment Rental & Leasing; Other Business Services; Aircraft Manufacturing; Management & Consulting Services; Motor Freight & Warehousing; Arrangement of Passenger Transportation.

COUNTY	EMPLOYMENT	SPECIALIZATION RATIO
Boone	19,969.44	2.21
Platte	8,794.11	2.15
Newton	4,108.89	1.86
Callaway	2,960.44	1.82
Adair	2,031.56	1.57
Phelps	2,758.33	1.45
Greene	23,509.11	1.39
Polk	1,169.67	1.37
St. Louis City	40,573.44	1.36
Johnson	2,267.56	1.33
Jasper	9,075.78	1.31
Butler	2,531.44	1.21
Nodaway	1,213.11	1.17
Cape Girardeau	5,479.11	1.15
St. Louis	83,444.89	1.08

Source: Covered Employment and Wages, Missouri Department of Economic Development.

V. Implications and Summary

Given that information technology (IT) is a targeted industry in Missouri, there is a need to delineate the entire IT value-chain within the state. One method to accomplish this is to determine and analyze where IT input and output clusters are located in Missouri. This information is significant in that decision-makers need to know where potential IT suppliers and consumers are located within the state. This allows businesses to better select facility locations, in that it identifies areas where IT suppliers and consumers are located. Also, it allows government officials to develop a strategy for recruiting IT firms, highlighting the state's existing supplier and consumer base. In general, the most optimal areas for the development of the IT industry are characterized by a sizable industry cluster employment base and high specialization.

In terms of backward-linkages, it appears that the IT industry purchases most of its inputs from the computer and data processing sector - indicating that the IT sector is highly dependent on other IT firms to provide needed inputs. In 2000, *First-Order Input* cluster jobs were located in metropolitan St. Louis, Kansas City, Columbia-Jefferson City, and Springfield. Additionally, jobs are also located in several rural areas of the state, including Barry County, Polk County and Wayne County. According to specialization ratios, 5 Missouri counties were highly specialized in First-Order Input cluster employment. These areas were located in suburban metropolitan areas and in several rural areas of the state. The most specialized counties in the state were Barry, Clay, St. Louis, Polk and Wayne.

The IT industry also purchases moderately from the wholesale trade and electronic components sectors to obtain various goods used in production; the communications sector to obtain telecommunications and internet services; and the real estate sector for facilities. In 2000, *Second-Order Input* cluster jobs were located in metropolitan St. Louis, Kansas City and Springfield. Additionally, jobs are also located in several rural areas of the state, including Cape Girardeau County, Cole County and Taney County. According to specialization ratios, 2 Missouri counties were highly specialized in Second-Order Input cluster employment. These areas were located in two rural areas of the state. The most specialized counties in the state were Dent and Putnam.

Lastly, the IT industry purchases sparingly from a host of business operation services in support of personnel, fiscal, legal and shipping operations. In 2000, *Third-Order Input* cluster jobs were located in metropolitan St. Louis, Kansas City, Springfield and Columbia. Additionally, jobs are also located in several rural areas of the state, including Cole County, Cape Girardeau County and St. Francois County. According to specialization ratios, 2 Missouri counties were highly specialized in Third-Order Input cluster employment. These areas were located in the south central area of the state. The most specialized counties in the state were Webster and Maries.

In terms of forward-linkages, it appears that the IT industry sells most of its outputs to the wholesale trade sector, who in turn sell IT goods to retailers for consumers to purchase. In 2000, *First-Order Output* cluster jobs were located in metropolitan St. Louis, Kansas City and Springfield. Additionally, jobs are also located in several rural areas of the state, including Cape Girardeau County, Cole County and Scott County. According to specialization ratios, 8 Missouri counties were highly specialized in First-Order Output cluster employment. These areas were located Kansas City and in several rural areas of the state. The most specialized counties in the state were Dent, Putnam, Clark, Chariton, Clay, Scott, Harrison and Dade.

Additionally, the IT industry also sells heavily to other IT firms, indicating that there is a high degree of interdependence among IT firms. In 2000, *Second-Order Output* cluster jobs were located in metropolitan St. Louis, Kansas City, Columbia-Jefferson City, and Springfield. Additionally, jobs are also located in several rural areas of the state, including Barry County, Polk County and Wayne County. According to specialization ratios, 5 Missouri counties were highly specialized in Second-Order Output cluster employment. These areas were located in the suburban metropolitan areas and in several rural areas of the state. The most specialized counties in the state were Barry, Clay, St. Louis, Polk and Wayne.

The IT industry also sells moderately to banks, the medical profession, and the telecommunications sector. Generally, the IT industry provides mission-critical services and software to these sectors that is specific to their business functions. In 2000, *Third-Order Output* cluster jobs were located in metropolitan St. Louis, Kansas City, Springfield and Columbia. Additionally, jobs are also located in several rural areas of the state, including Cape Girardeau County, Cole County and Butler County. According to specialization ratios, no Missouri counties were highly specialized in Third-Order Output cluster employment. However, Maries County and Jackson County had significant above average specialization.

Lastly, the IT industry sells sparingly to a host of professional service, transportation and manufacturing firms. The IT industry provides services and software to college and universities, hospitals, engineering and architectural firms, aircraft manufacturing and transportation firms, a variety of business and consulting services firms, and to the trucking and warehousing industry. Generally, these industries require a higher degree of IT infrastructure in order to operate.

References

Aldenderfer, Mark S. and Roger K. Blashfield. 1984. *Cluster Analysis*. Sage University Paper Series on Quantitative Applications in the Social Sciences, 44. London, UK: Sage Publications.

Bergman, Edward and Edward Feser. 1999. "Industry Clusters: A Methodology and Framework for Regional Development Policy in the U.S." In *Boosting Innovation: The Cluster Approach*, edited by T. Roelandt and P. den Hertog. Paris, FRANCE: Organisation for Economic Cooperation and Development.

Feser, Edward. 1998. "Old and New Theories of Industry Clusters." In *Clusters and Regional Specialisation: On Geography, Technology and Networks*, edited by M. Steiner. London, UK: Pion Publications.

Stough, Richard. 1998. *Technology in Virginia's Regions*. School of Public Policy Research Report. Fairfax, VA: George Mason University.

Appendix A - Statistical Methods

Distance Measures

Squared Euclidean Distance
This measure should be employed when using centroid, median or Ward's method of clustering. With this measure, the distance between two cases (*x* and *y*) is the sum of the squared differences between the values of the clustering variables.

$$S_{xy} = \sum_i (x_i - y_i)^2$$

Chebychev Distance
This measure should be employed when using centroid, median or Ward's method of clustering. With this measure, the distance between two cases is the maximum absolute difference between the values of the clustering variables.

$$S_{xy} = \max_i |x_i - y_i|$$

Clustering Methods

General Procedure

Begin with *N* clusters each containing one case. Denote the clusters 1 through *N*.

Where: *S* Matrix of similarity or dissimilarity measures
S_{ij} Similarity of dissimilarity measure between cluster *i* and cluster *j*.
N_i Number of case in cluster *i*.

- Find the most similar pair of clusters *p* and *q* (*p* > *q*). Denote this similarity S_{pq}. If a dissimilarity measure is used, large values indicate dissimilarity. If a similarity measure is used, small values indicate dissimilarity.
- Reduce the number of clusters by one through merger of clusters *p* and *q*. Label the new cluster *t (=q)* and update similarity matrix by the method specified to reflect revised similarities or dissimilarities between cluster t and all other clusters. Delete the row and column of *S* corresponding to cluster *p*.
- Perform the previous two steps until all entities are in one cluster.
- For each of the following methods, the similarity or dissimilarity matrix *S* is updated to reflect revised similarities or dissimilarities (S_{tr}) between the new cluster *t* and all other clusters *r* as given below.

Centroid Method

Update S_{tr} by:

$$S_{tr} = \frac{N_p}{N_p + N_q} S_{pr} + \frac{N_q}{N_p + N_q} S_{qr} - \frac{N_p N_q}{(N_p + N_q)^2} S_{pq}$$

Ward's Method

Update S_{tr} by:

$$S_{tr} = \frac{1}{(N_t + N_r)} \left[(N_r + N_p) S_{rp} + (N_r + N_q) S_{rq} - N_r S_{pq} \right]$$

Update the coefficient *W* by:

$$W = W + 0.5 S_{pq}$$

Note that for Ward's method, the coefficient given in the agglomeration schedule is really the within-cluster sum of squares at that step. For all other methods, this coefficient represents the distance at which the clusters *p* and *q* were joined.

ICA-0601-1

Analysis and reporting by David J. Peters, Planner.
Maps by Zachary Johnson, GIS Analyst.

June 2001

Direct all correspondence to David J. Peters:

580 Harry S. Truman Building
Research and Planning
Department of Economic Development
Jefferson City, MO 65102

TEL: (573) 522-2791
FAX: (573) 751=7385
E-MAIL: dpeters4@mail.state.mo.us
WEB: http://www.MissouriEconomy.org